Interconversion of Nonlocal Correlations

Nick S. Jones and Lluís Masanes
Department of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK

In this paper we study the correlations that arise when two separated parties perform measurements on systems they hold locally. We restrict ourselves to those correlations with which arbitrarily fast transmission of information is impossible. These correlations are called nonsignaling. We allow the measurements to be chosen from sets of an arbitrary size, but promise that each measurement has only two possible outcomes. We find the structure of this convex set of nonsignaling correlations by characterizing its extreme points. Taking an information-theoretic view, we prove that all of these extreme correlations are interconvertible. This suggests that the simplest extremal nonlocal distribution (called a PR box) might be the basic unit of nonlocality. We also show that this unit of nonlocality is sufficient to simulate all quantum states when measured with two outcome measurements.

I. INTRODUCTION

Measurements on parts of quantum states held by spatially separated parties cannot be used for superluminal signalling; in this respect quantum mechanics is a nonsignaling theory. John Bell [1] exposed a novel feature of the theory when he considered a gedanken experiment of the following form: two separated parties, Alice and Bob, locally measure two physical systems which were, at an earlier time, very close together. Bell found quantum states which display measurement outcome statistics which vary, as Alice and Bob change their measurements, in a way which cannot be explained by only assuming an exchange of classical information when the two systems were close together in the past. This behavior is termed quantum nonlocality and has partial experimental validation (for a discussion of experimental tests and loopholes see [2] and references therein).

Quantum mechanics is not the only conceivable theory that predicts correlations which, though they are nonsignaling, cannot be understood as having been established in the past. This paper investigates the structure of the set of all possible nonsignaling correlations and attempts to characterize these in information-theoretic terms.

Since quantum mechanics is so successful in its predictions, it might seem unusual to consider other theories, with different kinds of correlations, which are not physically instantiated. There are practical and foundational physical motives, as well as information-theoretic reasons, for considering a broader class of correlations.

Motivated by the technological promise of quantum information, there is a drive to understand the origins of quantum features which may have concrete applications. They could be direct consequences of the fact that quantum mechanics is a nonsignaling theory or, alternatively, exploit other features of the theory. Such concerns motivated the information-theoretic treatment of nonlocal correlations in [3]. A second reason for interest in general nonsignaling theories is foundational; given that quantum mechanics is a nonsignaling theory, what simplest possible extra features must be added to explain the results of our experiments? Popescu and Rohrlich [4] show that there exist nonsignaling correlations which cannot be reproduced by quantum mechanics: why is quantum mechanics not the most general kind of nonsignaling theory, what further constraints does it satisfy?

In the context of communication complexity and cryptography, interesting results have come from considering nonsignaling correlations. Van Dam [5] showed that, equipped with 'superstrong nonlocal' correlations, all bipartite communication complexity problems are rendered trivial (requiring only one bit of communication). There has also been work relating bit commitment to nonsignaling [6, 7, 8]. In cryptography, it is best to have security proofs that rely on a minimum number of principles; in [9], a key distribution scheme is presented which can be proved secure by only assuming nonsignaling.

Our work follows that of Barrett et al. [3]. They characterize the bipartite nonlocal correlations arising when Alice and Bob can perform one of two measurements, each with an arbitrary outcomes. They also provide results on the interconversion of correlations and consider the case of more than two parties.

In this paper, we consider the set of bipartite nonsignaling correlations, where each party performs one from an arbitrary set of measurements and each measurement has two possible outcomes (a reversal of the situation in [3]). This set is a convex polytope and we characterize it in terms of its extreme points. The structure of these extreme points has already been used by one of the authors [10] to show

that, for all nonsignaling theories, the more incompatible two observables are, the more uncertain their corresponding outcomes. From an information-theoretic perspective, we also prove that all nonlocal extremal correlations are interconvertible, in the sense that given sufficient copies, any one can simulate any other. Consequently, any nonlocal extremal distribution can simulate any non-extremal one. The simplest extremal nonlocal correlations are called Popescu-Rohrlich (PR) boxes [3]. One can thus consider the PR box as the unit resource of bipartite nonlocal correlations, in the same fashion as the singlet is considered the unit resource of quantum correlations. It is, as yet, unclear if they can serve as s sufficient unit in more general cases. Since quantum correlations are nonsignaling, all those within the polytope considered can be simulated by PR boxes. It has previously been shown [11] that all possible projective measurements on the singlet state of two qubits can be simulated using just one PR box and shared randomness (our result can be seen as an extension from projective measurements on singlets to POVMs on general bipartite quantum states).

This paper is structured in the following way. In Section II the set of nonsignaling correlations is characterized in terms of inequalities. Section III reviews past results and characterizes the structure of this set in terms of its extreme points. Section IV is devoted to the inter-convertibility of nonsignaling correlations. Section V concludes and shows, by giving an example, that the extreme points in more general cases have nonuniform marginals and thus lack the simple structures found in Section III and in [3].

II. NO-SIGNALLING CORRELATIONS

In what follows, we will consider two parties —Alice and Bob— each performing space-like separated operations. Each possesses a physical system which can be measured in several distinct ways and each measurement can yield several distinct results. Let x (y) denote the observable chosen by Alice (Bob) (these will also be called *inputs*), and a (b) be the result of Alice's (Bob's) measurement (these will also be called *outputs*). The statistics of these measurements define a joint probability distribution for the outputs, conditioned on the inputs, $P_{ab|xy}$, which satisfies the usual constraints:

$$P_{ab|xy} \geq 0 \;\; \forall\, a, b, x, y, \tag{1}$$

$$\sum_{a,b} P_{ab|xy} = 1 \;\; \forall\, x, y. \tag{2}$$

We consider the input x (y) to take values from an alphabet of length d_x (d_y), that is, $x \in \{0, ..., d_x - 1\}$ and $y \in \{0, ..., d_y - 1\}$. The output a (b) takes values from an alphabet of length d_a (d_b), $a \in \{0, ..., d_a - 1\}$ and $b \in \{0, ..., d_b - 1\}$.

A. No-signalling constraints

The requirement that Alice and Bob cannot signal to each other by using their correlations is equivalent to the condition that Alice's output is independent of Bob's input, $P_{a|x}$ is independent of y (and vice-versa):

$$\sum_{b} P_{ab|xy} = \sum_{b} P_{ab|xy'} \;\; \forall\, a, x, y, y', \tag{3}$$

$$\sum_{a} P_{ab|xy} = \sum_{a} P_{ab|x'y} \;\; \forall\, b, x, x', y. \tag{4}$$

For fixed d_x, d_y, d_a, and d_b, the set of probability distributions Eqs. (1, 2) is convex and has a finite number of extreme points. In other words, it is a convex polytope. It is known that the intersection of a polytope with an affine set, like the one defined by the no-signaling constraints (3, 4), defines another convex polytope. From now on, all distributions are assumed to belong to this set. In this paper such distributions are represented by tables of the form given in Table I.

y \ x	0	1	$\dots$	d_x-1
0	$P_{00\|00}$ $P_{10\|00}$ $P_{01\|00}$ $P_{11\|00}$	$P_{00\|10}$ $P_{10\|10}$ $P_{01\|10}$ $P_{11\|10}$		$P_{00\|d_x-1,0}$ $P_{10\|d_x-1,0}$ $P_{01\|d_x-1,0}$ $P_{11\|d_x-1,0}$
1	$P_{00\|01}$ $P_{10\|01}$ $P_{01\|01}$ $P_{11\|01}$	$P_{00\|11}$ $P_{10\|11}$ $P_{01\|11}$ $P_{11\|11}$		$P_{00\|d_x-1,1}$ $P_{10\|d_x-1,1}$ $P_{01\|d_x-1,1}$ $P_{11\|d_x-1,1}$
$\vdots$			$\ddots$	
d_y-1	$P_{00\|0,d_y-1}$ $P_{10\|0,d_y-1}$ $P_{01\|0,d_y-1}$ $P_{11\|0,d_y-1}$	$P_{00\|1,d_y-1}$ $P_{10\|1,d_y-1}$ $P_{01\|1,d_y-1}$ $P_{11\|1,d_y-1}$		$P_{00\|d_x-1,d_y-1}$ $P_{10\|d_x-1,d_y-1}$ $P_{01\|d_x-1,d_y-1}$ $P_{11\|d_x-1,d_y-1}$

TABLE I: This table represents a general probability distribution for d_x and d_y arbitrary and $d_a = d_b = 2$. The distribution is broken into $d_x \times d_y$ cells, with one cell for every input pair (x, y). Each cell specifies the probabilities of the four possible outcomes given these inputs (these must sum to one). The nonsignaling conditions require, for example, that $P_{00|00} + P_{01|00} = P_{00|01} + P_{01|01}$ and that $P_{01|00} + P_{11|00} = P_{01|d_x-1,0} + P_{11|d_x-1,0}$.

B. Local correlations

Local correlations are those that can be reproduced by parties equipped only with shared randomness. These are a proper subset of nonsignaling correlations. One can always write them as:

$$P_{ab|xy} = \sum_e p_e P_{a|xe} P_{b|ye}. \tag{5}$$

A protocol for generating $P_{ab|xy}$ is the following: With probability p_e Alice (Bob) samples from the distribution $P_{a|Xe}$ ($P_{b|Ye}$). It is known that the set of local correlations is a convex polytope with some of the facets being Bell-like inequalities [12]. The extreme points of this polytope correspond to local deterministic distributions, that is $P_{ab|xy} = \delta_{a,f(x)}\delta_{b,g(y)}$, where $f(x)$ and $g(y)$ map each input value to a single output value. Correlations that are not of the form (5) are called nonlocal.

C. Quantum Correlations

Quantum correlations are generated if Alice and Bob share quantum entanglement. These can be written as:

$$P_{ab|xy} = \mathrm{tr}[F_a^x \otimes F_b^y \rho], \tag{6}$$

where $\{F_0^x, ..., F_{d_a-1}^x\}$, $\{F_0^y, ..., F_{d_b-1}^y\}$, are positive operator valued measures for each x and y, and ρ is a density matrix. Though this set is convex, it is not a polytope; it includes all local correlations and also probability distributions which are nonlocal. It is, however, smaller than the full set of nonsignaling correlations. This was proved in [4] by providing an example of a nonsignaling distribution forbidden by quantum mechanics.

III. EXTREME NONSIGNALING CORRELATIONS

The full set of extremal distributions for the general situation where d_x, d_y, d_a and d_b is not yet characterized. In what follows, previous work considering the case where $d_a = d_b = d_x = d_y = 2$ and the case for $d_x = d_y = 2$, and both d_a and d_b arbitrary will be reviewed [3, 13]. Next, the extreme points for $d_a = d_b = 2$ and both d_x and d_y arbitrary will be presented.

A. Reversible local transformations

Applying reversible local transformations to a distribution does not change its nonlocal properties. We say that two distributions are *equivalent* if one can be transformed into the other by means of local

reversible transformations. Identifying *classes* of extremal distributions which are equivalent simplifies the task of categorizing all of them. It is sufficient to quote one representative element from each equivalence class. Let us list all possible local reversible transformations:

- Permute the ordered set of input values for each party, $(0, 1, \ldots d_x - 1)$ and $(0, 1, \ldots d_y - 1)$.
- Permute the ordered set of output values depending on the input. To indicate that a and b are associated with the particular inputs (x, y), the notation a_x and b_y will sometimes be used. Summarizing, one can apply a different permutation to each of the a_x (b_y), for each value of x (y).

B. Binary inputs and outputs

All extremal nonlocal distributions for the case $d_x = d_y = d_a = d_b = 2$ are equivalent to:

x / y	0		1	
0	1/2	0	1/2	0
	0	1/2	0	1/2
1	1/2	0	0	1/2
	0	1/2	1/2	0

(7)

(this format is explained in Table I) or alternatively:

$$p_{ab|xy} = \begin{cases} 1/2 & : \quad a + b \mod 2 = xy \\ 0 & : \quad \text{otherwise,} \end{cases} \tag{8}$$

where it is understood that a and b are locally uniformly distributed. This distribution is also called a PR box and constitutes the paradigm of nonlocality. PR boxes have their outputs together, depending on their inputs together; but they are nonsignaling since their outputs are (locally) random, obeying Eqs. (3,4).

C. Binary inputs and arbitrary outputs

Barrett et al. [3] provided the following characterization for the case where $d_x = d_y = 2$ and arbitrary d_a, d_b outcomes. Each inequivalent extremal nonlocal distribution is characterized by one value of the parameter $k \in \{2, \ldots \min(d_a, d_b)\}$. For each k, its corresponding distribution is

$$p_{ab|xy} = \begin{cases} 1/k & : \quad (b - a) \bmod k = xy \\ \\ 0 & : \quad \text{otherwise,} \end{cases} \tag{9}$$

where $a, b \in \{0, \ldots, k-1\}$ and are locally uniformly distributed. Note that Eq. (8) is recovered when $d_a = d_b = 2$.

D. Arbitrary inputs and binary outputs

In what follows, one of the main results of our paper is presented. We give a characterization of all extreme distributions for the case where d_x and d_y are arbitrary, and $d_a = d_b = 2$. The proof of this result is provided in the appendix.

Result 1: *Table II provides at least one representative element of all classes of extremal correlations for a given d_x and d_y. Each of these distributions is characterized as follows:*

1. *Giving two integers g_x and g_y, where $g_x \in \{2, 3, \cdots d_x\}$ and $g_y \in \{2, 3, \cdots d_y\}$ if the distribution is nonlocal, and $g_x = g_y = 0$ if the distribution is local.*

2. *And assigning perfect correlation or anti-correlation to all the cells with a question mark '?', that is*

$$\boxed{?\qquad} = \boxed{\begin{matrix}1/2 & 0\\ 0 & 1/2\end{matrix}} \text{ or } \boxed{\begin{matrix}0 & 1/2\\ 1/2 & 0\end{matrix}}. \tag{10}$$

x / y	0	1	2	$\ldots$	$g_x - 1$	g_x	$\ldots$	$d_x - 1$
0	$\begin{matrix}1/2 & 0\\ 0 & 1/2\end{matrix}$	$\begin{matrix}1/2 & 0\\ 0 & 1/2\end{matrix}$	$\begin{matrix}1/2 & 0\\ 0 & 1/2\end{matrix}$		$\begin{matrix}1/2 & 0\\ 0 & 1/2\end{matrix}$	$\begin{matrix}1/2 & 0\\ 1/2 & 0\end{matrix}$		$\begin{matrix}1/2 & 0\\ 1/2 & 0\end{matrix}$
1	$\begin{matrix}1/2 & 0\\ 0 & 1/2\end{matrix}$	$\begin{matrix}0 & 1/2\\ 1/2 & 0\end{matrix}$	?		?	$\begin{matrix}1/2 & 0\\ 1/2 & 0\end{matrix}$		$\begin{matrix}1/2 & 0\\ 1/2 & 0\end{matrix}$
2	$\begin{matrix}1/2 & 0\\ 0 & 1/2\end{matrix}$	?	?		?	$\begin{matrix}1/2 & 0\\ 1/2 & 0\end{matrix}$		$\begin{matrix}1/2 & 0\\ 1/2 & 0\end{matrix}$
$\vdots$				$\ddots$			$\ddots$	
$g_y - 1$	$\begin{matrix}1/2 & 0\\ 0 & 1/2\end{matrix}$	?	?		?	$\begin{matrix}1/2 & 0\\ 1/2 & 0\end{matrix}$		$\begin{matrix}1/2 & 0\\ 1/2 & 0\end{matrix}$
g_y	$\begin{matrix}1/2 & 1/2\\ 0 & 0\end{matrix}$	$\begin{matrix}1/2 & 1/2\\ 0 & 0\end{matrix}$	$\begin{matrix}1/2 & 1/2\\ 0 & 0\end{matrix}$		$\begin{matrix}1/2 & 1/2\\ 0 & 0\end{matrix}$	$\begin{matrix}1 & 0\\ 0 & 0\end{matrix}$		$\begin{matrix}1 & 0\\ 0 & 0\end{matrix}$
$\vdots$				$\ddots$			$\ddots$	
$d_y - 1$	$\begin{matrix}1/2 & 1/2\\ 0 & 0\end{matrix}$	$\begin{matrix}1/2 & 1/2\\ 0 & 0\end{matrix}$	$\begin{matrix}1/2 & 1/2\\ 0 & 0\end{matrix}$		$\begin{matrix}1/2 & 1/2\\ 0 & 0\end{matrix}$	$\begin{matrix}1 & 0\\ 0 & 0\end{matrix}$		$\begin{matrix}1 & 0\\ 0 & 0\end{matrix}$

TABLE II: This table gives a representative element of all classes of extreme points, where Alice (Bob) has d_x (d_y) different input settings, and g_x (g_y) of them are nondeterministic. Cells containing a '?' can either be perfectly correlated (like the cell corresponding to $x = y = 0$) or anti-correlated (like the cell corresponding to $x = y = 1$).

As one can see in Table II, each party has two kinds of input settings: (i) the deterministic ones ($x \geq g_x$ for Alice) have a fixed outcome, (ii) the nondeterministic ones ($x < g_x$ for Alice) have uniform probabilities for their corresponding outcomes, $P_{0|x} = P_{1|x} = 1/2$. There are g_x nondeterministic input settings and $d_x - g_x$ deterministic input settings in Alice's site and analogously for Bob. The representative distributions are chosen to have the outcomes for all deterministic input settings fixed to '0'.

The following observation will prove crucial. When the distribution is nonlocal, that is $g_x, g_y \geq 2$, there is always a PR box structure when both parties restrict to $x, y \in \{0, 1\}$.

Extreme points for which $g_x = d_x$ and $g_y = d_y$ can be algebraically characterized by: $a_x + b_y = \delta_{x,1}\delta_{y,1} + \sum_{(i,j)\in Q} \delta_{x,i}\delta_{y,j} \mod 2$. Here Q is any subset of the set $\{1, ..., d_x\} \times \{1, ..., d_y\} - \{(1, 1)\}$.

IV. INTERCONVERSION OF NONLOCAL CORRELATIONS

In this section we prove that all extremal nonlocal correlations with binary outputs can be interconverted. This means that all contain the same kind of nonlocality. By saying that the distribution $P_{ab|xy}$ can be converted into $P'_{ab|xy}$ we mean: given enough copies (realizations) of $P_{ab|xy}$, Alice and Bob can simulate the statistics of $P'_{ab|xy}$ for any value of x and y that they independently choose. We assume that the two parties can perform local operations and have unlimited shared randomness. This is a fair assumption because with these resources (shared randomness and local operations) we cannot create nonlocality.

Result 2: *All nonlocal extremal correlations with arbitrary d_x and d_y, and binary output ($d_a = d_b = 2$) are interconvertible.*

In order to prove this statement, we first argue that all extremal nonlocal correlations can simulate a PR box, and second, we prove that PR boxes are sufficient to simulate all extremal distributions. By recalling that all distributions can be written as probabilistic mixtures of extreme points (noting that such mixtures can be reproduced by shared randomness), one can also make the following statement:

Result 3: *PR boxes are sufficient to simulate all nonsignaling correlations with binary output ($d_a = d_b = 2$).*

By looking at Table II one can see that, if Alice and Bob share a nonlocal distribution ($g_x, g_y \geq 2$), they have a PR box when restricting $x, y \in \{0, 1\}$. This shows that a single copy of any nonlocal extremal distribution can simulate a PR box. Next, we present a protocol that allows Alice and Bob to simulate any distribution of the form described in Table II, by only using a finite number of PR boxes (8). This protocol is based on an idea presented in [5].

If $g_x = g_y = 0$ the distribution is local, and thus, it can be simulated with the protocol detailed in Section II.B without using PR boxes. When $g_x, g_y \geq 2$, however, such protocols cannot be used. Let us first describe how to make the simulation when Alice and Bob choose inputs $x \leq g_x - 1$ and $y \leq g_y - 1$. Within this range of input settings the outcomes a_x and b_y are locally random and they are either perfectly correlated ($a_x + b_y = 0 \mod 2$), or anti-correlated ($a_x + b_y = 1 \mod 2$). Equivalently, any distribution of the form defined by Table II for inputs $x \leq g_x - 1$ and $y \leq g_y - 1$ is equally well defined by a function:

$$F(x, y) = a_x + b_y. \tag{11}$$

Throughout this section all equalities are always modulo 2 and thus we omit the specification '(mod 2)'. Let us expand x and y in binary: $x = (x_1 x_2 \dots x_{n_x})$, $y = (y_1 y_2 \dots y_{n_y})$, where $n_x = \lceil \log_2 g_x \rceil$ and $n_y = \lceil \log_2 g_y \rceil$. The function $F(x, y)$ can always be expressed as a polynomial of the binary variables $x_1, \dots x_{n_x}, y_1, \dots y_{n_y}$. More specifically, one can always write $F(x, y)$ as a finite sum of products

$$F(x, y) = \sum_{i=1}^{2^{n_y}} P_i(x) Q_i(y), \tag{12}$$

where each $P_i(x)$ is a polynomial in the variables $\{x_1, x_2, ..., x_{nx}\}$, and each Q_i is a monomial in the variables $\{y_1, y_2, ..., y_{n_y}\}$. The sum has at most 2^{n_y} terms, because there are 2^{n_y} distinct monomials in the variables $\{y_1, y_2, ..., y_{n_y}\}$.

Let us describe the Protocol. Suppose Alice and Bob choose the input settings $x \leq g_x - 1$ and $y \leq g_y - 1$. Alice (Bob) evaluates the 2^{n_y} numbers $r_i = P_i(x)$ ($s_i = Q_i(y)$) depending on the x (y) chosen. Then, Alice (Bob) inputs the binary number r_i (s_i) in the i^{th} PR box and obtains the outcome a_i (b_i). They do such operations for all $i = 1, \dots 2^{n_y}$. Finally, each party computes its output of the simulated distribution (a_x, b_y) by summing the local outputs of the PR boxes:

$$a_x := \sum_{i=1}^{2^{n_y}} a_i \qquad\qquad b_y := \sum_{i=1}^{2^{n_y}} b_i. \tag{13}$$

The protocol works because of the next chain of equalities:

$$F(x, y) = \sum_{i=1}^{2^{n_y}} P_i(x) Q_i(y) = \sum_{i=1}^{2^{n_y}} r_i s_i = \sum_{i=1}^{2^{n_y}} (a_i + b_i) = \sum_{i=1}^{2^{n_y}} a_i + \sum_{j=1}^{2^{n_y}} b_j = a_x + b_y. \tag{14}$$

To see the third equality, just recall that for each PR box $a_i + b_i = r_i s_i$ holds.

Let us now consider the case where Alice picks an input $x \geq g_x$, she must then assign to a_x the corresponding deterministic value and analogously for Bob. One can see that the simulation protocol works for all values of x and y.

A corollary of Result 3 is the following. Since quantum correlations are nonsignaling, the statistics of any two-outcome measurements experiment on any bipartite quantum state, can also be simulated with PR-boxes (as noted in the introduction, this result extends [11]).

y \ x	0			1			2		
0	1/4	0	1/4	1/2	0	0	1/2	0	0
	0	1/4	0	0	1/4	0	0	1/4	0
	1/4	0	0	0	0	1/4	0	0	1/4
1	1/2	0	0	1/4	1/4	0	1/4	0	1/4
	0	1/4	0	1/4	0	0	1/4	0	0
	0	0	1/4	0	0	1/4	0	1/4	0
2	1/2	0	0	1/4	1/4	0	0	1/4	1/4
	0	1/4	0	1/4	0	0	1/4	0	0
	0	0	1/4	0	0	1/4	1/4	0	0

TABLE III: An extreme point of the nonsignaling polytope for 3 input settings each with 3 possible outcomes. Each cell contains 9 probabilities associated with the 3×3 possible outcome pairs.

V. DISCUSSION

In this paper we have given a complete characterization of the extremal nonsignaling bipartite probability distributions with binary outputs. We have grouped them into equivalence classes under local reversible transformations. One can see in Table II that these extremal distributions have a more complicated structure than in the binary input scenario (8,9). Nevertheless, if we consider purely nonlocal distributions ($g_x = d_x$ and $g_y = d_y$) all the marginals are unbiased ($P_{a|x} = P_{b|y} = 1/2$) and they are easily defined by specifying which input pairs (x, y) have correlated outputs and which (s, y) have anti-correlated outputs. In more general cases the extremal distributions stop showing these simple symmetries. An example of this more complex structure is given in Table III. This is an extremal distribution for the case $d_a = d_b = d_x = d_y = 3$, which we discovered numerically (that this is extremal can be verified by using arguments similar to those in Part 3 of the Appendix). Its corresponding marginals are not unbiased and there are also some input pairs, (x, y), for which, once the output of one party is fixed the outcomes of the other remain uncertain.

We have shown that all extremal nonlocal distributions with binary outputs are interconvertible. We have also given a specific protocol to implement this interconversion. By looking at the asymmetric structure of the extremal distribution in Table III, one sees that this protocol is not directly applicable in general. In particular it is an open question whether this distribution can be simulated by PR boxes. We conclude by noting that, just as treating the singlet as a unit of entanglement motivated numerous resource based questions (asymptotic interconversions, multipartite scenarios, etc) so too there is an analogous set of unanswered information-theoretic problems involving units of nonlocality.

Note added. After the completion of this work, the authors were made aware that similar results have been obtained by J. Barrett and S. Pironio [14].

Acknowledgments

The authors would like to thank J. Barrett, S. Popescu and T. Short for discussions. This work has been supported by the U.K. Engineering and Physical Sciences Research Council (IRC QIP).

APPENDIX

In this appendix we give the proof of Result 1. Firstly we show that any nonsignaling distribution can be expressed as a convex combination of distributions equivalent to ones of the form given in Table II. Secondly we show that all distributions of the form given in Table II are extremal.

Some simple definitions will prove useful throughout this appendix. The word 'cell' refers to the set of four outcome probabilities $P_{00|xy}$, $P_{10|xy}$, $P_{01|xy}$, $P_{11|xy}$ associated with the input pair (x, y). It will be useful to think of $P_{ab|xy}$ as a table of cells with d_x columns and d_y rows where, associated with each entry

of the table, (x,y), there is a cell of four probabilities (See Table I). We say that a cell has 'one zero' if it has at least one of the four entries it contains set to zero. We call $P_{a|x}, P_{b|y}\ \forall a,b,x,y$ the 'marginals'. Specifically we define $P_{a=0|x=i} \equiv l_i$ and $P_{b=0|y=i} \equiv m_i$. We now sketch the strategy adopted for Parts 1 and 2 of the proof.

An arbitrary distribution $P^{(1)}$ is expressed as a convex combination of two distributions:

$$P^{(1)} = \lambda_1 P_1^{(1)} + (1-\lambda_1) P_2^{(1)}. \tag{15}$$

We require that the new distributions, $P_1^{(1)}$ and $P_2^{(1)}$, have one more entry of their tables set equal to zero. Next we select one of them: $P_1^{(1)}$ or $P_2^{(1)}$. We then repeat the above decomposition for the selected distribution. A schematic of the approach is:

$$\begin{aligned}
P^{(1)} &= \lambda_1 P_1^{(1)} + (1-\lambda_1) P_2^{(1)}, && (16)\\
P^{(2)} &= P_{k_1}^{(1)}, \quad k_1 \in \{1,2\}, && (17)\\
P^{(2)} &= \lambda_2 P_1^{(2)} + (1-\lambda_2) P_2^{(2)}, && (18)\\
P^{(3)} &= P_{k_2}^{(2)}, \quad k_2 \in \{1,2\}, && (19)\\
P^{(3)} &= \lambda_3 P_1^{(3)} + (1-\lambda_3) P_2^{(3)}, && (20)\\
&\vdots\\
P^{(F)} &= \lambda_F P_1^{(F)} + (1-\lambda_F) P_2^{(F)}, && (21)
\end{aligned}$$

where the $P^{(i)}$ are probability distributions $P_{ab|xy}$ expressed as vectors and $\lambda_i \in [0,1]$. At each step a distribution with one more entry set to zero is selected. It may happen that a distribution $P^{(i)}$ will already have a zero at the position demanded in the next step (e.g. if $P^{(i)}$ is already extremal). The expression $k \in \{1,2\}$ (e.g. in Eqs. (17,19)) indicates that the consecutive steps of the proof hold independently of which of the two distribution is chosen. The procedure stops when the new distribution chosen, $P_{k_F}^{(F)}$, is equivalent to one of the form given in Table II. We will see that this procedure, based on successive zeroing of entries, always finishes.

The proof of Result 1 is in three parts.

- In Part 1 we show that any probability distribution can be expressed as a convex combination of probability distributions which have at least one zero in every cell and which have the same marginals ($P_{a|x}$, $P_{b|y}$, $\forall\, a,b,x,y$) as the original distribution.
- In Part 2 we show that any probability distribution with one zero in every cell can be expressed as convex combinations of probability distributions which are locally equivalent to Table II.
- In Part 3 we show that all distributions of the form defined in Table II are extremal.

Part 1

This Part is broken in two. We first show that every cell can be expressed as a convex combination of two cells which satisfy the following constraints. Each has the same marginals as the original cell and also has at least one of their four entries set to zero. The nonsignaling conditions (3,4) mean that all cells in the same column, x, have the same marginals $P_{a=0|x} = l_x$, $P_{a=1|x} = 1 - l_x$. Consider a cell with marginals $l_x \geq m_y \geq \frac{1}{2}$. Given l_x and m_y, there is one free parameter, c, needed to completely specify the cell (x,y):

$$\boxed{\begin{array}{ll} c & m_y - c \\ & \\ l_x - c & 1 + c - m_y - l_x \end{array}}_{\,xy.} \tag{22}$$

By the above notation we mean: $P_{00|xy} = c$; $P_{10|xy} = m_y - c$; $P_{01|xy} = l_x - c$; $P_{11|xy} = 1 + c - m_y - l_x$.

If we require the positivity of the four elements of the cell, then $c \in [m_y + l_x - 1, m_y]$. One can readily check that all cells with allowed values of c can be written as convex combinations of the two cells where $c = m_y$ and $c = m_y + l_x - 1$:

$$\boxed{\begin{array}{cc} c & m_y - c \\ \\ l_x - c & 1 + c - m_y - l_x \end{array}}_{xy} = \lambda \boxed{\begin{array}{cc} m_y & 0 \\ \\ l_x - m_y & 1 - l_x \end{array}}_{xy} + (1-\lambda) \boxed{\begin{array}{cc} m_y + l_x - 1 & 1 - l_x \\ \\ 1 - m_y & 0 \end{array}}_{xy.} \tag{23}$$

Instead of $l_x \geq m_y \geq \frac{1}{2}$, cells can satisfy different inequalities e.g. $m_y \geq l_x \geq \frac{1}{2}$ or $m_y \geq \frac{1}{2} \geq l_x$. Using symmetries, one can see that, whatever inequalities are satisfied, any cell can be expressed as a convex combination of two, one zero, cells in a similar manner.

In the second half of this Part we generalise from single cells to the whole distribution. An iterative procedure of the form described in Eqs. (16-21) can be applied. Any distribution, $P^{(1)}$, can be expressed as a convex combination of two distributions, $P_1^{(1)}$ and $P_2^{(1)}$. Both $P_1^{(1)}$ and $P_2^{(1)}$ have all cells equal to $P^{(1)}$, except for the cell $(x = 0, y = 0)$. This cell has the same marginals as in $P^{(1)}$ but also has one more zero. $P_1^{(1)}$ (or $P_2^{(1)}$) can again be expressed as a convex combination of two distributions $P_1^{(2)}$ and $P_2^{(2)}$ which are identical to $P_1^{(1)}$ (or $P_2^{(1)}$) - they also have $(x = 0, y = 0)$ as a one zero cell - except that they also both have $(x = 0, y = 1)$ as a one zero cell (with the same marginals). This procedure can be extended to all cells until the final step has a probability distribution which is a convex combination of two distributions which have at least one zero in every cell. It follows that any probability distribution can be expressed as a convex combination of probability distributions which have at least one zero in every cell and which have the same marginals as the original distribution.

Part 2

In this part we show that distributions with one zero in every cell can be expressed as a convex combination of distributions equivalent to Table II. The argument exploits the fact that the only parameters describing distributions with one zero in every cell are their marginals. It considers first a cell (1), then a column (2) and finally a generic table (3).

1. A Cell

In this section we identify the constraints on the marginals in one zero cells. Consider a cell in the first column $(x = 0, y = i)$ with the form:

$$\boxed{\begin{array}{cc} m_i & 0 \\ \\ l_0 - m_i & 1 - l_0 \end{array}}_{0i.} \tag{24}$$

Note that the marginals are $P_{a=0|x=0} = l_0$, $P_{b=0|y=i} = m_i$. Positivity requires that $l_0 \in [m_i, 1]$. For the same marginals, l_0 and m_i, if $P_{00|0i} = 0$, instead of $P_{10|0i} = 0$:

$$\boxed{\begin{array}{cc} 0 & m_i \\ \\ l_0 & 1 - l_0 - m_i \end{array}}_{0i,} \tag{25}$$

then $l_0 \in [0, 1 - m_i]$. If $P_{01|0i} = 0$ instead then $l_0 \in [0, m_i]$. Finally, if $P_{11|0i} = 0$ then $l_0 \in [1 - m_i, 1]$. In an arbitrary one zero cell, l_0 will thus lie in one of four ranges:

$$[0, 1 - m_i], \qquad [0, m_i], \qquad [m_i, 1], \qquad [1 - m_i, 1], \tag{26}$$

depending on which of its four elements is zero. Part of the information in Eq. (26) can be expressed as follows. We call v_1 the lower bound on l_0 and the upper bound v_2 ($l_0 \in [v_1, v_2]$). Without knowing which of the four entries is zero in the cell, or even knowing the value of m_i, we do know from Eq. (26) that $v_1 \in \{0, 1 - m_i, m_i\}$ and $v_2 \in \{1, 1 - m_i, m_i\}$. This observation will be used in the ensuing subsection.

2. A Column

In the following we use the constraints on l_0 deduced in the preceding section (Eq. 26) to express a column of a distribution's table as a convex combination of two simpler columns. Recall that all cells in column x (row y) have the same marginal $P_{a|x}$ ($P_{b|y}$) by Eqs. (3,4). Each probability distribution has d_y cells in the column $x=0$. If there is one zero in every cell of the column, there will be d_y overlapping ranges (see Eq. (26)) in which l_0 can lie (while keeping all other marginals, m_i, constant). It is possible that l_0 will be uniquely determined by these ranges (e.g. if cell $(0,1)$ requires $l_0 \in [0, m_1]$ and cell $(1,2)$ requires $l_0 \in [m_2, 1]$ and $m_1 = m_2$). One knows that there is at least one value of l_0 consistent with all ranges, but generically, l_0 will lie in a range of the form $l_0 \in [u_1^{(1)}, u_2^{(1)}]$. Here $u_1^{(1)}$ is the largest lower bound on l_0 and $u_2^{(1)}$ the smallest upper bound, with $u_1^{(1)} \in \{0, 1-m_0, 1-m_1 ..., m_0, m_1 ...\}$ and $u_2^{(1)} \in \{1, 1-m_0, 1-m_1 ..., m_0, m_1 ...\}$. An arbitrary distribution, $P^{(1)}$, with marginal $P^{(1)}_{a=0|x=0} = l_0$ will have $l_0 \in [u_1^{(1)}, u_2^{(1)}]$ ($u_1^{(1)}$, $u_2^{(1)}$ as defined previously). One can check that it can always be expressed as a convex combination of two distributions $P_1^{(1)}$ and $P_2^{(1)}$ with $P^{(1)}_{1\ a=0|x=0} = l_0 = u_1^{(1)}$ and $P^{(1)}_{2\ a=0|x=0} = l_0 = u_2^{(1)}$ respectively.

Example:

$$\begin{array}{|cc|}\hline m_0 & 0 \\ l_0-m_0 & 1-l_0 \\ \hline 0 & m_1 \\ l_0 & 1-l_0-m_1 \\ \hline l_0 & m_2-l_0 \\ 0 & 1-m_2 \\ \hline \end{array} = \lambda \begin{array}{|cc|}\hline m_0 & 0 \\ 0 & 1-m_0 \\ \hline 0 & m_1 \\ m_0 & 1-m_0-m_1 \\ \hline m_0 & m_2-m_0 \\ 0 & 1-m_2 \\ \hline \end{array} + (1-\lambda) \begin{array}{|cc|}\hline m_0 & 0 \\ m_2-m_0 & 1-m_2 \\ \hline 0 & m_1 \\ m_2 & 1-m_2-m_1 \\ \hline m_2 & 0 \\ 0 & 1-m_2 \\ \hline \end{array} \tag{27}$$

Above is a simple example of the procedure described. Without knowing the specific values of m_0, m_1 and m_2, and without even looking where the zeros are in each cell of the column, we do have the basic knowledge that $l_0 \in [u_1^{(1)}, u_2^{(1)}]$ where $u_1^{(1)} \in \{0, 1-m_0, 1-m_1, 1-m_2, m_0, m_1, m_2\}$ and $u_2^{(1)} \in \{1, 1-m_0, 1-m_1, 1-m_2, m_0, m_1, m_2\}$. By looking at this specific case we can now refine our bounds on l_0. In the following we suppose, as an example, that $1-m_1 > m_2$. From the cell $(0,0)$ on the left hand side of Eq. (27) we know, by positivity, that $l_0 \in [m_0, 1]$. From the cell $(0,1)$ we know that $l_0 \in [0, 1-m_1]$. From the cell $(0,2)$ we know that $l_0 \in [0, m_2]$. Taking the largest lower bound and the smallest upper bound from these ranges, and recalling that $1-m_1 > m_2$, one finds that $l_0 \in [m_0, m_2]$. The left hand side of Eq. (27) can be expressed as a convex combination of two columns where $l_0 = m_0$ and $l_0 = m_2$ and the m_i are kept constant. Note that each of the two columns on the right hand side of Eq. (27) contains a cell which has two zeros. These two columns each have one more zero than the column on the left hand side.

3. Generalizing

In this subsection we provide a procedure which shows that any distribution with one zero in every cell can be expressed as convex combinations of probability distributions which are locally equivalent to Table II. We first provide the loop of the procedure and second the condition for its termination. This approach is effectively a generalization of the decomposition of the column given in the preceding section. From Part 1 it is sufficient to consider only distributions which have one zero in every cell.

Loop: The loop considered is of the form described in Eqs. (16-21): (I) A distribution is expressed as a convex combination of two simpler distributions (II) one of these distributions is selected and then becomes the distribution in step (I) - the loop then continues.

In what follows we will follow the loop through two cycles.

(I) A starting distribution $P^{(1)}$ will have l_0 constrained to lie in a range $l_0 \in [u_1^{(1)}, u_2^{(1)}]$ where $u_1^{(1)} \in \{0, 1-m_0, 1-m_1..., m_0, m_1...\}$ and $u_2^{(1)} \in \{1, 1-m_0, 1-m_1..., m_0, m_1...\}$. It can be expressed as a convex combination of two distributions, $P_1^{(1)}$ and $P_2^{(1)}$. These satisfy the further constraints that $P^{(1)}_{1\ a=0|x=0} = l_0 = u_1^{(1)}$ and $P^{(1)}_{2\ a=0|x=0} = l_0 = u_2^{(1)}$ respectively. (This implies that $P_1^{(1)}$ and $P_2^{(1)}$ each have one cell which has two zeros in their $x = 0$ columns.)

(II) The distribution $P_{k_1}^{(1)}$, $k_1 \in \{1, 2\}$ is chosen as $P^{(2)}$.

(I) $P^{(2)}$ will have $l_0 = u_1^{(1)}$ or $u_2^{(1)}$. There is now one less parameter in the table because two of the marginals have been related. $l_0 = u_{k_1}^{(1)}$ will also be constrained to lie in a new range $l_0 \in [u_1^{(2)}, u_2^{(2)}]$ where $u_1^{(2)} \in \{0, 1-m_0, 1-m_1..., m_0, m_1...1-l_0, 1-l_1..., l_0, l_1...\}$ and $u_2^{(2)} \in \{1, 1-m_0, 1-m_1..., m_0, m_1...1-l_0, 1-l_1..., l_0, l_1...\}$. $P^{(2)}$ can be written as a convex combination of a distribution $P_1^{(2)}$, with $l_0 = u_{k_1}^{(1)} = u_1^{(2)}$, and $P_2^{(2)}$ with $l_0 = u_{k_1}^{(1)} = u_2^{(2)}$.

(II) The distribution $P_{k_2}^{(2)}$, $k_2 \in \{1, 2\}$ is chosen as $P^{(3)}$.

(I) ...

Depending on the choices made at each step (II) the procedure creates distributions satisfying a chain of equivalences between their marginals:

$$l_0 = u_{k_1}^{(1)} = u_{k_2}^{(2)} = u_{k_3}^{(3)} = ... = u_{k_F}^{(F)}, \tag{28}$$

which will be specified by the string $(k_1, k_2, ...k_F)$ with $k_i \in \{1, 2\}$. The nature of $u_{k_F}^{(F)}$ will be discussed as part of the termination conditions. Noting which sets the $u_1^{(i)}$ and $u_2^{(i)}$ are chosen from, a chain of equivalences could, for example, be of the form $l_0 = m_2 = 1 - m_6 = l_1 = ...$. Note that after each cycle the distributions have one less parameter as more and more of their marginals are related to each other. The procedure shrinks the number of free parameters as it converges towards extreme points (these have no free parameters).

The first equivalence in a chain of equalities can only be $l_0 = m_n$ or $l_0 = 1 - m_n$ for some n (the $l_0 = 0$ or 1 case will be discussed as part of the termination conditions). This is explained by noting that in the first cycle l_0 is constrained by cells in the same column (see the preceding subsection). It follows that $u_1^{(1)}$ lies in the set $\{0, 1-m_0, 1-m_1..., m_0, m_1...\}$ and $u_2^{(1)}$ lies in $\{1, 1-m_0, 1-m_1..., m_0, m_1...\}$ which only depend on the values of the m_i. After the first cycle in which $l_0 = m_n$ or $l_0 = 1 - m_n$, the cells in both column $(x = 0)$ and the row $(y = n)$ will provide constraints on l_0. This is because the cells in row $(y = n)$ all depend on m_n. With some thought, one sees that in general $u_1^{(j)}$ will thus lie in the set $\{0, 1-m_0, 1-m_1..., m_0, m_1...1-l_0, 1-l_1..., l_0, l_1...\}$ and $u_2^{(j)}$ in $\{1, 1-m_0, 1-m_1..., m_0, m_1...1-l_0, 1-l_1..., l_0, l_1...\}$ and these depend on both m_i and l_i.

Termination conditions: We now discuss loop termination. It terminates, after F cycles, in two distinct ways.

(a) When $u_{k_F}^{(F)} = 0$ or 1

(b) When $u_{k_F}^{(F)} = 1 - u_{k_g}^{(g)}$ for $g < F$. This is only satisfied if $u_{k_F}^{(F)} = 1/2$.

An example of case (b) would be $l_0 = m_2 = 1 - m_6 = l_1 = ... = 1 - m_2$, which implies that all of these numbers must be $1/2$.

After the loop terminates, several marginals from the set of all l_i and m_i will have been set to either $0, 1$, or $1/2$ (the procedure as a whole always terminates, as there are a finite number of marginals to be equated). If there exists a set of marginals which have not been fixed to one of these values, a new marginal l_i (or m_i) from this set can be chosen. The form of the above loop can then be repeated by studying constraints on this new variable.

By repeating this procedure, all marginals, m_i and l_i, will be absorbed into a chain of equalities terminating in $0, 1$ or $1/2$. A probability distribution equivalent to Table II will be the only possible result. It will generally be necessary to perform some local relabelling to obtain distributions of the form of Table II. For example, the outcomes for all deterministic input settings have to be fixed to '0'.

Part 3

The following proves by contradiction that all distributions of the form defined in Table II are extremal. Suppose that a particular distribution $P_1^{(F)}$ of form defined in Table II is not extremal. It can thus be expressed as a convex combination of more than one distribution. Positivity requires that these distributions have a zero where $P_1^{(F)}$ has a zero.

Suppose, from Table II, that $P_1^{(F)}$ has $g_x = g_y = 0$ then all of its cells have three zeros. This distribution cannot be expressed as a convex combination of two distinct distributions with the same zeros, since normalization fixes the fourth entry of each cell to be one. $P_1^{(F)}$ is the only distribution with these zeros.

If $P_1^{(F)}$ has $g_x, g_y \geq 2$ it will have some cells with three zeros (if $g_x < d_x$ and $g_y < d_y$) and some with two zeros. As noted above, the three zero cells have their fourth entry fixed by normalization. A study of the distribution of zeros in the four cells (i, j), $i, j \in \{0, 1\}$ shows that all remaining non-zero entries are forced to be one-half. $P_1^{(F)}$ is the only distribution with its particular distribution of zeros.

---

[1] J. S. Bell, Physics (Long Island City, N.Y.) **1**, 195 (1964).
[2] D.G. Collins, Ph.D. thesis, University of Bristol, Department of Physics (2002), available at http://rogers.phy.bris.ac.uk/theses.html.
[3] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu and D. Roberts, Phys. Rev. A **71**, 022101 (2005).
[4] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[5] W. van Dam, e-print quant-ph/0501159.
[6] S. Wolf and J. Wullschleger, e-print quant-ph/0502030.
[7] T. Short, N. Gisin and S. Popescu, e-print quant-ph/0504134.
[8] H. Buhrman, M. Christandl, F.Unger, S. Wehner and A. Winter, e-print quant-ph/0504133.
[9] J. Barrett, L. Hardy, and A. Kent, e-print quant-ph/0405101.
[10] Ll. Masanes, A. Acín and N. Gisin; *General properties of no-signaling theories*, in preparation.
[11] N. J. Cerf, N. Gisin, S. Massar and S. Popescu, Phys. Rev. Lett. **94**, 220403 (2005).
[12] A. Peres, Found. Phys. **29**, 589 (1999).
[13] B.S. Tsirelson, Hadronic J. Suppl. **8**, 329 (1993).
[14] J. Barrett and S. Pironio, e-print quant-ph/0506180.